This paper presents first steps towards the development of a formal model of the research process. We evaluate the use of simulation as a tool for the evaluation of research strategies in nascent research organizations faced with the absence of significant data. We start by modeling the research process by using the "Publish or Perish" paradigm, a well-known criteria of evaluation of research. We demonstrate the use of this model for researchers to evaluate the effects of selection of a particular publishing venue over time. We then perform various experiments using this basic idea. By means of various visualization techniques, we see how researchers with similar publishing policies might self-organize in the form of groups. We also evaluate the effects of giving higher weights to articles in journals and see where the effects of publishing in these venues breaks even for both top as well as average acceptance rates.

1 INTRODUCTION

Research is, arguably, one of the key processes, by which the human civilization has achieved its ultimate glory. It is through research that we are now at the stage where we can differentiate ourselves clearly from other life forms on our planet. Looking at research from another perspective, even modeling and analysis of all aspects of a single human being can be fairly hard, if not impossible; analysis of the processes that lead to the now ongoing and streamlined process of research is bound to be extremely complex. This is the complex area that we attempt to touch upon in this paper by formally providing a first model suitable to the current workflow of research and publishing processes.

We analyze the research process from the perspective of nascent research organizations attempting to eventually compete with existing research bodies with the goal of achieving this as soon as possible. Such organizations are faced with a challenge which their peers never faced; for them it was a plain natural evolutionary process (as shall be discussed in the background section). Their challenges include evaluation of policies to ensure that their research output eventually matches or excels peer organizations in countries with a well developed networks of institutes. And this has to happen in as short a time as possible. In the absence of real data to make intelligent decisions, simulation offers an effective tool for analyzing policy decisions alongside a validation mechanism for comparison with real data as it starts pouring in, with the passage of time.

We start by examining how employees of academic research institutions i.e. researchers, have to decide upon venues to maximize their research output in the upcoming years. However, with such a wide variety of publishing venues and acceptance rates as well as reviewing times, individual researchers might have to either select the best possible option at the current time or else follow the organization’s policy of publishing. The policy can range from publishing in the same country in terms of conferences or else publishing mostly in highly cited journals. In well-developed research organizations, at any given time, there might be researchers at all levels which can provide useful guidance on these matters. However, for researchers in nascent organizations, even such seemingly simple decisions can be fairly hard to get at logically. The goal of this work is to develop a well-defined model and provide a set of simulation experiments to assist researchers in answering these questions according to their particular domain.
1.1 Structure of the paper

The rest of the paper is structured as following: First, we start by talking about the growth of a handful of prominent research institutes from history. We then develop a model for a research institute. Next, various simulation experiments are conducted and results are discussed. Finally, we conclude the paper with a discussion of our currently ongoing and future explorations.

2 BACKGROUND

Research, what is it? And how can we assist in the growth of nascent research institutes in developing nations. Ray Paul (Paul 2007) refers to research as “. . . original investigation undertaken in order to gain knowledge and understanding.”

Although modeling and simulation of research is a new area, for this paper we have covered an initial literature survey in two related areas and subsequently introduced concepts from other work as and when it is referred:

1. Background about modeling humans and processes.
2. Background about research organizations.

2.1 Modeling and Simulating Humans

The use of modeling and simulation is very popular in domains where significant historic data is not available. In (An, Jeng, Lee, and Ren 2007), An et. al. use systems dynamics modeling and simulation to help in the effective workforce management life cycle. Whereas in (Siebers, Aickelin, Celia, and Clegg 2007), Siebers et. al. use intelligent agents to understand management practices and workplace productivity. In (Ferrin, Miller, and McBroom 2007), Ferrin et. al. use simulation to maximize hospital and emergency room throughput. Social norms have been modeled using multiagent systems in (Hexmoor, Venkata, and Hayes 2006).

2.2 Historical Growth Of Research

Being very abstract, we start the examination of the research process by first examining the growth history of a few of the world-renowned universities which are known for their clear research graduate and post-doctoral bias listed here in an alphabetical order:

2.2.1 Caltech

According to (Caltech 2008): “The mission of the California Institute of Technology is to expand human knowledge and benefit society through research integrated with education. We investigate the most challenging, fundamental problems in science and technology in a singularly collegial, interdisciplinary atmosphere, while educating outstanding students to become creative members of society.” As of December 2006, the faculty included 293 Professorial faculty, 104 Emeriti, 66 Research faculty. 49 other faculty, 87 visiting faculty, 534 Postdoctoral scholars, 24 Senior postdoctoral scholars and 115 visitors. One noticeable difference from Cambridge is that (Prize 2008) mentions Caltech’s beginnings as rooted in a modest little college founded in Pasadena in 1891 by wealthy former abolitionist and Chicago politician Amos Throop. Initially named Throop University, the school changed its name to Throop Polytechnic Institute in 1893. In its first fifteen years, Throop served the local community, teaching a great variety of subjects, from arts and crafts to zoology, with considerable emphasis on vocational training. Also, mentions Caltech’s history to be divided into two distinct eras: “The first Caltech era was created by Hale, Millikan, and Noyes. Thirty years later, after World War II, the physicists Lee Alvin DuBridge and Robert Bacher did the job all over again. DuBridge, the head of MIT’s wartime radar project, became Caltech’s new president in 1946. Bacher, the leader of the Los Alamos atomic bomb project’s “G” Division (the “G” stood for gadgets), arrived in 1949 to head up the division of physics, mathematics, and astronomy and later became the Institute’s first provost.”

2.2.2 University of Cambridge

According to (University 2008): “The University of Cambridge is one of the oldest universities in the world and one of the largest in the United Kingdom. Its reputation for outstanding academic achievement is known world-wide and reflects the intellectual achievement of its students, as well as the world-class original research carried out by the staff of the University and the Colleges. Many of the University’s customs and unusual terminology can be traced to roots in the early years of the University’s long history.” Cambridge university was established in the town of Cambridge and started as a religious school. In the early days, the University had no premises of its own relied on churches as sites for its public ceremonies. Lectures, disputations and lodgings were found in private houses which frequently changed hands or went out of use (University 2008).

2.2.3 Oxford University

According to Wikipedia (Wikipedia 2008b), the University of Oxford or Oxford University is the oldest university in the English-speaking world. The university traces its roots back to at least the end of the 11th century. Oxford predates Cambridge and historically, it is considered that Cambridge was formed by some dissenting scholars from Oxford. As such, like Cambridge, it has religious roots.
2.2.4 MIT

The plan for the MIT started when William Barton Rogers drafted a plan for a scientific school, calling it a “Plan for a Polytechnic School in Boston” as in (Library 2008). In 1861, MIT was incorporated with first classes in Boston in 1865. By 1869, the first laboratory was established at MIT.

3 MODEL DEVELOPMENT

Simplifying at various levels, the current research and reporting process can be modeled as shown in the 1.

The research workflow starts when a researcher or a group of researchers start working individually or collectively on one or more seemingly innovative ideas. The ideas are to be correlated with existing literature to understand the complex of the state of the art. The researcher attempts modeling, experimentation and simulation, continuously re-defining criteria of success as experience is gained. Success may come very early or else never. In our model, we consider success as a research output which is believed to be both innovative and significant by the researcher, herself. However, like other beliefs, this may or may not be factual so does not guarantee an actual publication in any venue. Based on peer-review, there are both pros and cons to this process. In peer-review, one benefit is the thorough evaluation by existing authorities or peers to be reasonable enough to be published. On the other hand, peer-review might also result in a collective inertia where even factual but radical ideas might receive strong opposition from the group and might be denied publication.

The researchers can always choose to not report it. Why would a researcher not want to publish their results? Being the outcome of human deliberation and thoughts, the reason for not attempting to publish could range from anything as diverse as results not being significant enough yet to or maybe better reportable along with other results or else plain laziness. So, to ensure we make a formal model, we only consider results as success if they are considered to be worth publishing by the researcher.

Next, the researchers embark on the task of writing the results in the form of a report. This would typically be based on the format guidelines specified by the targeted Conference or Journal. Afterwards, this report may then be submitted. Once the report gets in the hands of the editorial board/committee of the conference or journal, it ends up one way or the other into the hands of a group of reviewers.

The reviewers give their reviews and depending upon a large number of factors including consensus, eventually the report results in either an acceptance for publication or else is rejected, hopefully with a set of reviews offering guidance to the authors. In case of a rejection, the researcher is at least armed with these reviewer’s comments, so she can start to work again on the same project and perform more experimentation, generate better results, and the process starts again.

3.1 Temporal Modeling

All of these above-mentioned tasks take time. Although it is, in actuality, a very complex workflow, we can however conceptually model the time it could take from start to end. So, let us find the time it could take to publish. Let us suppose that the time for the modeling, experimentation (in the form of simulation or actual testing) and the evaluation of literature survey i.e. the entire process, is the time of working \( t_{work} \). Whereas, the time to write, format and edit the report is the reporting time \( t_{rpt} \). After the paper is submitted to the publication venue, the time it would take during the peer-review is the review time \( t_{rvw} \). Assuming it is accepted, the time it takes to actually be published (assuming if the venue supports such publishing) for online would be \( t_{oe} \) and to be available in print \( t_{prt} \). \( t_{pub} \) is the time to publish (When the work becomes accessible to other researchers) and would be minimum of online or print publishing. So, essentially the total time it took to perform research can be written as following.

\[
t_{rsh} = t_{work} + t_{rpt} + t_{rvw} + t_{pub}
\]

Where the time to publishing would be as following:

\[
t_{pub} = \min(t_{oe}, t_{prt})
\]

3.2 Decision Modeling

There are decisions to be made at every step of the process. As an example, from the point of view of researchers:

- Should we research alone or should we research in collaboration with students or other researchers?
- Are these results really innovative or is there any other work, which I have missed?
- Should I submit this report to a conference or to a journal?
- Which conference or journal is this article suitable for?

From the management perspective, decisions can again be of various types. Some examples are given below:

- Should we fund this work?
- Is this researcher sufficiently productive or not?
- For researchers new to the field (just after an MS or a PhD), should we encourage research group formation from the start as maybe some of these researchers may be less productive if they work alone?
• Or should we let the researchers try it on their own and let natural self-organization take place as the researchers discover how they work best?

Other questions could be from the Journal Editor’s perspective:

• Does minimizing the reviewing time really have any impact on the popularity of the journal?
• How can we encourage authors to write for us?
• Would aiming for a better ranking really interest authors or else actually deter authors by giving the impression of selectivity?

3.3 Probability Modeling

At each stage, we have different tasks, and as researchers become more experienced, they might gradually increase the chance of getting their work to a level acceptable for publishing. This can be modeled using probabilities. So, as an example, during experimentation, we have the probability of success $P_{Scs}$. Afterwards, there is also a probability that a researcher gets accepted in a particular venue. The factors involved in this are innumerable so it is easy to get lost and say that we cannot even model and simulate some of them. However, if we can simplify and focus on the average standard acceptance rates and assume that all researchers submit their work after completion, we can actually get some reasonably interesting results (as shown shown in subsequent sections).

3.4 Modeling Publishing venues

It is standard practice in academia that for researchers to progress, they must follow the “Publish or Perish” paradigm (Wikipedia 2008a). The two major publishing venues are conferences and journals. Opinions on publishing in either of them vary significantly. On one hand, Journals with higher impact factors are typically considered better venues for research. However on the other hand, they have certain problems associated with them.

As an example, for Journals, we have the following benefits:

• The typical chance of being cited is higher for archival Journals.
• Management typically values archival journals more than conferences, even if they are archival conferences.
• Once an article has been accepted for publication, even if it is not yet in print or online, researchers already get the “paper benefit” of being able to cite the work at least themselves or else put it on the resume for benefits from the organization.
Published articles might have a longer shelf life. However, with their benefits, there are some problems as well:

- Journals have a much longer time of review. The typical reason for this can range from the typical relatively longer articles and the review times of return by the referees.
- Even when an article has been accepted, unless the journal supports online access, articles can have significantly long time to publish. This means that other authors will not be able to evaluate and cite this work during this waiting period. Online first publishing, adopted by some publishers, however offers a somewhat effective solution to this problem.
- By the time articles have been published, the actual state of the research may be a lot more advanced so this can result in the “light year effect” similar to astronomy where what we are viewing is actually not the state of the art but something which has more of a historical significance.
- The references in the article may be quite old and outdated by then as well.

Unlike journals, publishing in conferences have some additional set of pros and cons such as:

- It typically takes more of a funding to publish and present in a conference. (Charges can include travel as well as the registration expenses)
- Good conference can however be a good way of networking.
- Occasionally, some conference paper may be publishable in the form of an extended version with significant changes, in a particular Journal as well, if the journal allows it.

Our goal here is to make decisions about the selection of conferences and Journals easier and informed rather than randomly. As such, we start with a basic model of researchers. These are classified into three main types:

1. Researchers which like to publish only in conferences.
2. Researchers which like to publish only in journals.
3. Researchers which like to publish in conferences as well as journals.

4 VERIFICATION AND VALIDATION

The model was verified using a number of techniques mentioned in texts such as (Banks, II, Nelson, and Nicol 2005). Some of the techniques used included making a flow diagram, making the operational model self-documenting as well as animated and having an Interactive Run Controller (IRC).

Because of the complexity of this domain, there is an absence of real data. As such, we used techniques such as testing the model for face validity and a variation of the Turing test validation (Banks, II, Nelson, and Nicol 2005) where we compared the results with the norms of the research fields such as preferences of journals. In addition, being a completely unexplored topic, here we attempt to answer possible questions about modeling and simulation of the research process.

Q: Can we even model research?
Ans: Research is a complex process. However, research is a human trait and humans are themselves part of a complex system. Social simulation has been successfully applied to simulating human, animal and artificial systems.

Q: What good would modeling research do?
Ans: In the developed world, the process of development of research institutes has taken several hundred years. In today’s world, if a country wants to initiate research and develop similarly producing academic organizations, she may not have the benefit of time on her side. Modeling can at the very least show effects of policy when there is not sufficient amount of data to make an intelligent decision.

Q: How can we assume a certain acceptance rate or review times?
Ans: Acceptance rates vary a lot. We have demonstrated a basic simulation of acceptance rates based on data from Computer Science domain. However, this simulation shows that other researchers can easily vary these figures for their particular domains. AAAI is a top AI conference and its acceptance rate varies from 17 to 30 percent from 1986 to 2008 (AAAI 2008). In addition, top Journals get more good papers. As an example JMLR, a relatively high impact CS Journal (5.952 in 2004) has a 27% acceptance rate (JMLR 2008). Other comparisons give similar acceptance rates. Another set of comparisons of conference and journal acceptance rates for the domain of Computer Science is given in (Science and Board 1994). Other acceptance rates give Journals much higher values but the actual figure could be considered lower because highly cited journals actually get better articles typically (Online 2008).

5 SIMULATION

We have used NetLogo (Wilensky 1999) to develop our simulation experiments. Our model is based on actual conference and journal review and published times obtained from the websites of various reputable conferences and journals. Our model is based on a visual representation of researchers. Each researcher has a publishing policy. This policy is either to publish solely in conferences, or to publish solely in journals or to publish in either of these.
We also model conference and journals by means of an average acceptance rate and a reviewing time. So, each researcher attempts to work on research and then tries to publish in their preferred venue. This venue takes the required average review time of either \( t_c \) for conference review times or else \( t_j \) for journal review times. Then, subsequently based on the acceptance rates of \( ra_c \) and \( ra_j \) for conferences and journals respectively, the paper are accepted. The simulation was developed to visually summarize the growth of researchers. So, researchers which succeed in publishing more than others are able to move to the top faster than the rest. We can see the end results of a simulation experiment in Figure 2.

Extensive simulations were performed on this model. The idea was to look for patterns in different ways of evaluating publishing venues. Because of the stringent longer reviewing time for journals, in one set of experiments, we also chose different weights given to journal articles.

### 5.1 Simulation constants and variables

We have based our simulation constants on acceptance figures from the domain of Computer Science journals and conferences; these values are described as following:

- Top conference (average) acceptance rate was chosen as average of AAAI acceptance rates. 
  \( ra_{tc} = 24\% \)
- Average acceptance rate for Computer Science Conferences was calculated from the data given in (Online 2008) to be \( ra_{ac} = 36.72\% \) based on around 600 events from the 1981 to 2008.
- Average reviewing time for conferences was chosen as \( t_{rvw} = 1.5 \) months
- Top Journal acceptance rate was chosen as \( ra_{a,j} = 10\% \)
- Average journal acceptance rate was chosen as \( ra_{a,j} = 30\% \)
- Average reviewing time for journals was chosen as \( t_{rvw} = 6 \) months
- Our experiments performed simulations with professors ranging from 100 to 1000 or 2000 professors, depending on experiments.
- Number of repetitions per experiments: 3

In terms of notation however, please note the following abbreviations are being used:

1. NOP stands for No of professors
2. TCP stands for Total Conference preferers Publications
3. TJP stands for Total Journal preferers Publications
4. TNP stands for Total No preference professors’ Publications

### 5.2 Experiment 1

The first set of basic simulation experiments was conducted using average acceptance rates for both conferences and journals. X-Axis shows the effects of varying number of professors while the y-Axis show the number of publications. The series are for the three type of professors. The results are plotted as given in the Figure 3.

![Figure 3: Publications in average Conference and Journals Acceptance](image)

We then conducted a similar simulation experiment with top conference and journal acceptance rates. The results are shown graphically in the Figure 4.

![Figure 4: Publications in top Conference and Journals Acceptance](image)

### 5.3 Experiment 2

In the second set of experiments, we evaluated the effect of varying the journal impact. The idea is that occasionally, the impression is that some universities prefer one archival
journal article to several conference papers. So, to see the results of this effective publications, we performed comparisons of how the researchers preferring the respective types of venues would fare in such circumstances.

The first experiment takes average journal and conference acceptance rates. The graphical representation of these results is shown in Figure 5.

Figure 5: Varying Impact of Journals (Avg acceptance)

The second experiment takes top journal and conference acceptance rates. The graphical representation of these results is shown in Figure 6.

Figure 6: Varying Impact of Journals (Top acceptance)

5.4 Experiment 3

The next set of experiments is on varying the journal review times. First we vary the journal review times taking into consideration average acceptance rates for conferences and journals.

Graphically, we see this in the fig 7
On the other hand, if we take the top acceptance rates for journals and conferences, we get the following data.
Graphically, we see this in the fig 8

6 DISCUSSION OF RESULTS

In the set of simulation experiments, some interesting results can be observed. First of all, it is interesting to see the model resulting in a detailed study of variables which can be tweaked to give different types of results for different types of publishing venues. The three types of experiments are discussed next:

1. In the first set of basic experiments, it appears that for acceptance rates in the average Computer Science venues, it is best to publish in conferences
and people with no preferences at around 1.4 times weight to each journal article.

4. For the second experiment set with top conference and journal rates, we see journal preferring researchers to come out better when each journal article is ranked at around 2.3 conference articles.

5. In the third set of experiments, we vary the reviewing time of journals from three months to an year. First, we examine the average acceptance rates and see that varying review time does not seem to impact the journal preferring people assuming journal articles are given same weight as conference papers.

6. In case of top conferences and journals, we again see that varying the review time does not impact the journal preferring researchers much.

6.1 Emergence of Self-Organization

In the visual simulation experiments, the researchers with similar policies join self-organize together to form groups in terms of their level of expertise. This emergence of self-organization in groups might be responsible for peer group formation in real research.

7 CONCLUSION & FUTURE WORK

In this paper, we have presented an analysis of the research process. We have also presented a first version of a model of the process. We have demonstrated how agent-based modeling and simulation can be used to evaluate various policies for publishing in journals and conferences. We have examined the problem of choosing conference versus journals as publishing venues. In different scenarios, our results demonstrate that occasionally groups may form based on similar policies in publications. We are currently working on developing a model for adaptation in research which is based on how researchers learn from others. We have also not considered the effects of citations in the current paper and have instead used top conferences and journals from social norms and impact factors as an indicative of the same. In future work however, we would also like to evaluate the effects of citations directly and simulate how they effect publications in addition to formation of naturally self-organized research groups.

REFERENCES


AUTHOR BIOGRAPHIES

MUAZ NIAZI obtained his BS EE (Honors) from Pakistan and his MS in Computer Sciences from Boston University in 1996 and 2004 respectively. He is currently working on his research at Foundation University, Pakistan as well as being a Doctoral student at the University of Stirling, UK. His current areas of research interest are self-organization in Peer-to-peer, social and ambient intelligent Systems. His email address is <thhgttg@gmail.com>.

DR. AMIR HUSSAIN obtained his BEng (1st Honours) and PhD both from the University of Strathclyde in Glasgow in 1992 and 1996 respectively. He is currently a Reader in Computing Science at the University of Stirling. His principal research area of interest and expertise is novel interdisciplinary research for intelligent modeling and control of complex systems. Dr. Hussain has over 100 publications in numerous international journals, books and refereed international conference proceedings. Dr. Hussain is Editor-in-chief of (Springers) Journal of Cognitive Computation, serves on the Advisory/Editorial Boards of four other international journals and has been Plenary speaker at various international Conferences and Workshops. He is Chairman of the IEEE UK & Ireland Chapter of the Industry Applications Society and a technical Consultant for Industrial Systems and Control (ISC) Ltd. He is founding co-chair of the biennial Conference series: Brain Inspired Cognitive Systems (BICS), and the IEEE International Conference on Engineering of Intelligent Systems (IEEE ICEIS). He is a Senior Member of the IEEE and a Fellow of the UK Higher Education Academy. His email address is <ahu@cs.stir.ac.uk>

DR. ABDUL RAUF BAIG (PhD 2000, University of Rennes II, France), is currently Professor of Computer Science at National University of Computer and Emerging Sciences, Pakistan. His research interests include solving games, data mining and applicative optimization problems with computational intelligence techniques. His email address is <rauf.baig@nu.edu.pk>

DR. SAEED AKHTAR BHATTI is a Professor and Head of Department of Software Engineering at Foundation University. He has an MSc. from Bradford University (UK) and a Ph.D. in Computer Sciences from ABU (Nigeria). He has authored several books in the domain of Computer Science as well as Simulation and Modeling. His email address is <saeedbhatti2000@yahoo.com>.